Social Norms, Local Interaction, and Neighborhood Planning

Matthew Haag* and Roger Lagunoff†

July 23, 1999‡

Abstract

This paper examines optimal social linkage when each individual's repeated interaction with each of his neighbors creates spillovers. Individuals differ across rates of time preference. A planner must choose a local interaction system or *neighborhood design* before observing the realization of these rates. Given the planner's choice of design and a realization of discount factors, each individual plays a repeated Prisoner's Dilemma game with his neighbors. We introduce the concept of a *local trigger strategy equilibrium (LTSE)* to describe a stationary sequential equilibrium in which, for any realization of discount factors, each individual conditions his cooperation on the cooperation of at least one "acceptable" group of neighbors. The presence of impatient types implies that some free riding may be tolerated in equilibrium. When residents' discount factors are known to the planner, the optimal design exhibits a cooperative "core" and an uncooperative "fringe." Uncooperative (impatient) types are connected to cooperative ones who tolerate their free riding so that social conflict is kept to a minimum. By contrast, when residents' discount factors are independently distributed, the optimal design partitions individuals into maximally connected *cliques* (e.g., cul-de-sacs). In that case, each person's cooperation decision becomes a pure local public good. Finally, if types are correlated, then incomplete graphs with small overlap (e.g., grids) are possible.

JEL Fields: C7, D62, D7.

Keywords: repeated games, local interaction, social norms, neighborhood design, local trigger strategy

---

*Department of Economics, Georgetown University, Washington DC 20057 USA. Phone: (202) 687-6704. e-mail: haagm@gusun.georgetown.edu.

†Department of Economics, Georgetown University, Washington DC 20057 USA. Phone: (202) 687-1510. e-mail: lagunofr@gunet.georgetown.edu. Web site: http://www.georgetown.edu/lagunoff/lagunoff.htm

‡We thank seminar participants at Columbia, Georgetown, Penn, the Midwest Theory Conference, and the 1999 North American Summer Econometric Society Meetings for useful comments and suggestions.

1 Introduction

It is well known that repeated play can mitigate free rider problems when social spillovers exist. Standard results establish that cooperative outcomes are attainable when individuals are sufficiently patient. These results usually pertain to environments with global interactions — all individuals interact with one another in each repetition of the stage game.[1]

Less is known when interactions are local.[2] Yet, in many instances local rather than global interactions prevail. For example, a neighbor who leaves his porch light on illuminates houses in direct view, but not those located further down the street; an office worker may disrupt the work of those in adjascent cubicles, but not in non-adjascent ones; a gas station that lowers its price attracts customers away from the station across the street but, perhaps, not from the one two blocks away.

These local interactions settings share the characteristic that each person interacts with only a subset of the relevant population. Moreover, the interaction need not be transitive; one's strategic "neighbors" may not be the same as one's "neighbor's neighbors." This paper examines repeated play in such settings. Specifically, our goal is to compare consequences of repeated play across different spatial designs. It turns out that some designs are more conducive than others to socially desirable outcomes. We therefore address the question of the *optimal spatial or neighborhood design* when free rider problems are localized.

A Simple Example

To see how the interaction design itself makes a difference, consider a repeated game in a neighborhood with three individuals. We first suppose that these three individuals live in houses arranged in a *cul-de-sac*. A cul-de-sac is a circular street in which each of the houses is in plain view of the other two. The importance of this, for our purposes, is that (1) the actions of each neighbor are consequential to both of the others, and (2) these actions are observable to the others. In Figure 1 below, the cul-de-sac is represented abstractly as a graph; the nodes are the houses, and the links express both the presence of potential spillovers and the information flows.

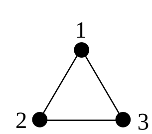


Figure 1:

[1]See, for example, Fudenberg and Maskin (1986). Some results also pertain to environments with random, pairwise matching. See Kandori (1991), Ellison (1994), and Okuno-Fujiwara and Postlewaite (1995).

[2]Related models are discussed in Section 5.

Each individual can choose one of two actions, "C" or "D". Action "C" corresponds to "cooperative" behavior, such as mowing the lawn, taking in one's garbage can at the end of the day, leaving an outdoor light on, or volunteering to maintain a communal garden. Conversely, action "D" is defined as the "deviant" or "uncooperative" action. In each of these examples, an individual's cooperative action confers external rewards to others. The payoffs resulting from the interaction between any pair of neighbors is given by the Prisoner's Dilemma game in Figure 2 below.

	C	D
C	c, c	-ℓ, d
D	d,-ℓ	$0, 0$

Figure 2

In Figure 2, assume $d > c > 0 > -\ell$. Payoff c is the gain from mutual cooperation, d the gain from "deviant" behavior, and ℓ the loss from being cheated. We also assume that $2c > d - \ell > 0$ so that mutual cooperation is Pareto undominated by payoffs in the convex hull of the payoff set,[3] and the aggregate payoff from having some cooperation is preferable to having none at all.

While Figure 2 describes the payoffs from a single interaction between two individuals, we examine activities, such as those mentioned above, in which an individual cannot isolate his behavior toward one neighbor from his behavior toward other neighbors. In each period, Neighbor 1 chooses an action, either "C" or "D", that simultaneously affects each of his neighbors. This is indicated in Figure 3 below.

Neighbor 1: C ↗

Neighbor 2	C	D
C	c, c	-ℓ, d

Neighbor 3	C	D
C	c, c	-ℓ, d

Neighbor 1: D ↘

Neighbor 2	C	D
D	d,-ℓ	$0, 0$

Neighbor 3	C	D
D	d,-ℓ	$0, 0$

Figure 3

[3]Any payoff in convex hull can be arbitrarily approximated by a time averaged pure behavior strategy.

If, for example, all three choose "C", then Neighbor 1 receives a payoff of $2c$. Given this setup, the interaction is essentially global. Hence, a standard analysis of repeated games may be used to determine bounds on the discount factor δ_i for each individual $i = 1, 2, 3$ under which full cooperation, i.e., all neighbors choose "C", is an equilibrium of this repeated interaction game. To find this bound, we examine a perfect equilibrium in which the "grim trigger strategy" — defection is met with permanent reversion to the Nash equilibrium — is used. If other neighbors adopt this strategy, then each individual i will choose "C" if $c > (1 - \delta_i)d$, i.e., the payoff from permanent cooperation outweighs the one shot gain from defection. Full cooperation is therefore an equilibrium if

$$\delta_i \geq \frac{d-c}{d},\ i = 1, 2, 3. \tag{1}$$

We now compare this to an alternative neighborhood design. Suppose instead that the three individuals live on a grid. In Figure 4 below, Neighbor 1 is flanked by each of the other two. The difference between this and the cul-de-sac is that there are no spillovers between Neighbors 2 and 3. Their interaction is limited to Neighbor 1, who interacts with each as before. We assume further that each neighbor's information coincides with the spillover linkages.[4]

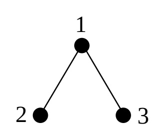


Figure 4:

As displayed above, the one dimensional grid is isomorphic to the triangle of Figure 1 absent one of the links — the one between 2 and 3. As with the cul-de-sac, Inequality (1) is necessary in order for full cooperation to be a perfect equilibrium. However, unlike the cul-de-sac, (1) is not sufficient. Perfection entails that Neighbor 1 be able to credibly punish either 2 or 3 for defecting. When the grim trigger punishments are used, this "perfection" constraint requires that Neighbor 1 satisfy $(1 - \delta_1)d \geq (1 - \delta_1)(c - \ell) + \delta_1(1 - \delta_1)d$. For instance, if Neighbor 3 defects to "D", then Neighbor 1 must prefer to retaliate by choosing "D" himself, rather than tolerate 3's behavior, even for just one period.[5] Rewriting this inequality yields an upper bound constraint,

$$\delta_1 \leqslant \frac{d-c}{d} + \frac{\ell}{d}. \tag{2}$$

[4]If, for instance, a neighbor enjoys viewing another neighbor's flower bed, then the spillover coincides with his observations of his neighbor.

[5]Standard arguments show that it suffices to check for single period deviations.

Interestingly, if $\frac{d-c}{d} + \frac{\ell}{d} < 1$ then an excessively patient Neighbor 1 with $\delta_1 > \frac{d-c}{d} + \frac{\ell}{d}$ will not punish a single deviator on either side. Instead, he will tolerate uncooperative behavior by one of his two neighbors. In such a case, potential gain to "containment" of the bad behavior on one side outweighs the gain to retaliation when a neighbor chooses "D". Therefore, while the constraint in (2) describes the incentive to credibly *punish* an out-of-equilibrium deviation, the opposite of (2) describes the in-equilibrium incentive to *tolerate* one uncooperative neighbor.

Inequalities (1) and (2) define three relevant intervals of discount factors. Low types satisfy $\delta_L < \frac{d-c}{d}$, and hence, will never cooperate. "Tolerant" types satisfy $\delta_T \geq \frac{d-c}{d} + \frac{\ell}{d}$, meaning that they are patient enough, not only to cooperate themselves, but will also tolerate uncooperative behavior from one other. Finally, "intolerant" types are cooperative but intolerant of uncooperative behavior, i.e., $\frac{d-c}{d} \leqslant \delta_I < \frac{d-c}{d} + \frac{\ell}{d}$. Various combinations of three possible discount factors, δ_L, δ_I, and δ_T can be used to illustrate two important points about the design problem.

1. In a local interaction environment, patient individuals cannot necessarily obtain full cooperation as a (sequential) equilibrium outcome.

Suppose, for instance, that all individuals are exceedingly patient, i.e., $\delta_i = \delta_T,\ i = 1, 2, 3$. Now compare the two different neighborhood designs of Figures 1 and 4 assuming the maximum possible degree of equilibrium cooperation in each. It is easy to see that a full-cooperation equilibrium exists in the cul-de-sac in Figure 1. However, because (2) is violated for Neighbor 1, the full-cooperation equilibrium does not exist for the grid of Figure 4. The reason is that Neighbor 3 can choose "D" with impunity (alternatively, if Neighbor 3 cooperates then Neighbor 2 can choose "D" with impunity). Since Neighbor 1 will tolerate such a defection, Neighbor 3 has no reason to continue to cooperate.

Therefore, the cul-de-sac has, in this sense, higher aggregate welfare than the grid. One might be tempted to argue that this is generally the case, since the former allows credible retaliation that deters bad behavior. Such retaliation is not possible in the grid. However, other considerations may reverse this social ordering.

2. A design with incomplete connectivity may be socially preferred to one with maximal connectivity.

Consider: $\delta_1 = \delta_T,\quad \delta_2 = \delta_I,\quad \delta_3 = \delta_L$. Notice that unlike the previous example, full cooperation is not possible in any spatial arrangement. Specifically, Neighbor 3 will never choose "C". However, there is a partial cooperation equilibrium on the street grid in Figure 4. Neighbor 3 chooses "D", Neighbor 1 chooses "C", thereby tolerating Neighbor 3's bad behavior, and Neighbor 2 choose "C".

As before, we now compare this to the cul-de-sac. It turns out that with these discount factors, the only equilibrium in the cul-de-sac is one in which all individuals choose "D". The reason is straightforward. Neighbor 3's discount factor falls below the cooperation threshold so that he always finds "D" profitable. While this behavior is tolerable to Neighbor 1, it is intolerable to Neighbor 2. Neighbor 2's discount factor satisfies both (1) and (2) which means that he is patient enough to cooperate in equilibrium, but not patient enough to tolerate a neighbor choosing "D". Therefore, in a cul-de-sac Neighbor 2 will only respond to 3's choice by choosing "D", leaving Neighbor 1 no alternative but to choose "D".[6] In this case, then, the grid has higher aggregate welfare than the cul-de-sac. The former allows the more tolerant, cooperative types to shield behavior of uncooperative types from the less tolerant ones.

An Overview of the Model and Results

The example highlights an important tradeoff. If individuals are likely to be cooperative types, then structures with a high degree of social interaction, such as large cul-de-sacs, are preferable. However, in the presence of uncooperative types, a social planner may prefer to design structures that limit the social interaction among neighbors. This may entail smaller cul-de-sacs (e.g., 2-person rather than 3-person in the example) or buffered structures such as the grid in Figure 4.

Which structure dominates depends on the distribution over rates of time preference. Hence, we cannot use standard analysis of repeated games even when a Folk Theorem applies (as it does in the cul-de-sac).[7] Moreover, we do not suppose that a planner would necessarily know the discount factors (types) of the individuals. We, therefore, introduce heterogeneity *ex post* by having the planner choose a design that "works well on average" when discount factors are drawn from a population wide distribution.

Section 2 describes a benchmark model which formalizes the problem faced by a planner. A planner must choose a local interaction structure before observing the realization of each resident's discount factor. We refer to such a structure as a *neighborhood design*. A neighborhood design is an undirected graph where the links determine whether any two individuals interact. Information flows are assumed to coincide with the links in a neighborhood design.

To make the optimal choice, the planner anticipates how individuals will interact with their neighbors in a repeated Prisoner's Dilemma game once the spatial plan is set and types are realized. Rather than study the entire, complicated set of sequential equilibrium interactions, we limit attention to a particular subclass. We examine norms of conduct that rely on simple punishments to enforce social cooperation. The concept of a *local trigger strategy equilibrium (LTSE)* is introduced to describe a stationary sequential equilibrium in which each individual chooses to condition his cooperation on the continued cooperation

[6]More generally, one can show that if $c > d - \ell$ then (2) implies nonexistence of any sequential equilibrium with a positive aggregate payoff in the cul-de-sac.

[7]With some notable exceptions (see Harrington (1989) and Lehrer and Pauzner (1999)), the individuals in repeated game models are assumed homogeneous with respect to rates of time preference.

of at least one "acceptable" group of neighbors. For instance, in the grid Neighbor 1 may choose to cooperate whenever any one of his two neighbors continues to do so. LTSE admit possibly some degree of "in-equilibrium" free riding. We specifically focus on optimal LTSE — those that maximize the planner's criterion in a given neighborhood design. The set of LTSE is defined and characterized in Section 3.

Results on optimal neighborhood designs are contained in Section 4. The first of our main results characterizes the "full information" solution when the distribution over discount factors is degenerate. With full information, the optimal neighborhood design exhibits a cooperative "core" and uncooperative "fringe." Sufficiently patient neighbors ("cooperative types") are maximally connected to one another. However, impatient or "uncooperative" types are also connected to certain individuals in the cooperative group who are able to tolerate some free riding. Those who are both cooperative yet comparatively intolerant of free riders are connected to fewer of them. In this sense, the optimal design minimizes the degree of social conflict.

We then consider the case where types are iid. With iid types, the planner cannot prevent social friction *ex post*, and so neighborhood size is the primary instrument used to limit interaction when frictions are likely *ex ante*. In this case, the optimal solution is shown to partition individuals into identical, maximally connected graphs, or *cliques*. Since individuals in each clique are maximally connected to each other but are disconnected to those in other cliques, the behavior of each individual constitutes a local public good for the group. The optimal clique size varies depending on the payoff and distributional characteristics.

Optimal designs with low neighbor overlap (e.g., grids) are possible, but correlation in the joint distribution over types is required. We construct examples where grid designs are optimal when an exchangeable distribution draws from a mixture of both uncooperative types and cooperative types of various degrees of tolerance.

Section 5 concludes with a review of related literature and a discussion of extensions and modifications. Finally, Section 6 is an appendix with proofs of the results.

2 The Model

Consider a society with infinitely lived individuals denoted by the finite set M with $m = |M|$. Time is discrete and indexed by $t = 0, 1, \ldots,$. This society faces an environment with network externalities. At the beginning of time $t = 0$, a residential planner chooses a local interaction system that determines who interacts with whom. We refer to this structure as a *neighborhood design* which is described as a collection of subsets $N = (N_1, \ldots, N_m)$. For each individual i, the set N_i is the collection of individuals with whom i interacts each period. We refer to this subset as "i's neighborhood". Hence, $j \in N_i$ means that j is i's neighbor. We assume that the relation is symmetric so that $j \in N_i$ iff $i \in N_j$. We will sometimes use an alternative notation more common in the literature on local interaction systems: $\langle M, \sim \rangle$ where $\sim$ is a symmetric, irreflexive binary relation on $M \times M$. We write $j \sim i$ when i and j are neighbors.

The neighborhood design simultaneously describes both a pattern of network externalities and the informational flows in this society. That is, $i \in N_j$ means not only that i and j are neighbors, but also that i and j each observe the history of the other's behavior.[8] Some examples are given below. Figure 5a exhibits four residents connected as a part of a 1-dimensional lattice. Each individual interacts only with adjacent individuals. Figure 5b describes a maximally connected graph or *clique*. All neighbors are connected to each other. This is the connectivity that might result from individuals living in a cul-de-sac. Finally, Figure 5c exhibits two connected clusters joined by a "buffer" individual. This buffer has the ability to insulate the behavioral effects of one neighborhood cluster from another.

Figure 5a

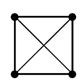


Figure 5b

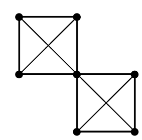


Figure 5c

For simplicity, it is assumed that each period, each resident can take one of two possible actions. Let $a_i^t \in \{C, D\}$ denote the action taken by individual i in period t. Action "C" connotes the "cooperative" action while "D" connotes "uncooperative" or "deviant" behavior. Individual i's payoff resulting from the interaction of each of his neighbors is summarized by the symmetric Prisoner's Dilemma game in Figure 2.

Let $a_{N_i}^t = (a_i^t, (a_j^t)_{j \in N_i})$ denote the behavior of everyone in i's neighborhood, including i himself. Individual i's t-period *personal history* is defined as the list $h_i^t = (\, a_{N_i}^1, a_{N_i}^2, \ldots, a_{N_i}^{t-1} \,)$. A standard notation denotes H_i as the set of all t-period personal histories for $i \in M$.

Individual i's dynamic payoff is the discounted sum $\sum_t (1 - \delta_i) \delta_i^t u_i(a_{N_i}^t)$ where δ_i is i's discount factor, and u_i is i's stage payoff function. Individual i's payoff at date t is the sum of the payoffs from each interaction. Hence, if k individuals in neighborhood N_i play "D" at time t then i's temporal payoff is $(n_i - k)c - k\ell$ if he plays "C", and is $(n_i - k)d$ if he plays "D". This specification captures both negative scale effects of congestion if most neighbors choose "D", and scale effects of positive spillovers if most neighbors choose "C".

We assume that at date $t = 0$, before play begins, each individual's discount factor is randomly determined. Let G denote the joint distribution over vectors $\delta = (\delta_1, \ldots, \delta_m)$. The marginal distribution over i's discount factor is denoted by G_i.

Once a vector $\delta = (\delta_1, \ldots, \delta_m)$ is realized, it is commonly known to all individuals at

[8] See Section 5 for extensions that break this link.

the start of play. Each individual then chooses a behavior strategy f_i in the ensuing local interaction game. For each personal history h_i, we write $f_i(h; \delta) \in \{C, D\}$ to denote i's action as a function of history h given a parameter vector of realized discount factors δ.[9] Societal behavior in the interaction game can now be completely described by a profile $f = (f_i)_{i \in N}$ of strategies. Clearly, the "uncooperative" profile f in which everyone plays "D" regardless of history is a sequential equilibrium.

The model differs from standard repeated game models in two respects. First, the neighbor relation obviously need not be transitive. That is, one's neighbor's neighbor need not be one's neighbor. Hence, each individual plays a repeated game with a possibly distinct subset of the population. Second, the dispersion of discount factors δ_i introduces heterogeneity across individuals. The presence of any number of impatient individuals implies that the Folk Theorem will not generally apply. Moreover, since strategic behavior is type-contingent, certain neighborhoods may tolerate some degree of in-equilibrium "cheating."

A planner who knows G must choose a neighborhood design before the realization δ. A useful analogy is to that of a city planner who knows something about the aggregate population characteristics, but does not know specific identities of the residents who will live in the houses she plans to build. If, following the planner's chosen neighborhood design $N = (N_1, \ldots, N_m)$, a strategy profile f is anticipated, then the planner's criterion is

$$E\left[\frac{1}{m}\sum_{i \in M}\sum_{t}(1-\delta_i)\delta_i^t u_i(\tilde{a}^t_{N_i}(f))\right], \tag{3}$$

where E is the expectation operator taken with respect to distribution G, and $\tilde{a}^t_{N_i}(f)$ is the action profile of i's neighbors induced by anticipated profile f. The maximum posible value of (3) is $(m-1)c$ which occurs when everyone permanently cooperates and every pair of individuals are linked.[10]

Unfortunately, two issues arise that render (3) inadequate for our purposes. First, if the planner does not directly influence the choice of f, then any design can maximize (3) provided that the residents play the uncooperative equilibrium following any other chosen design.[11] To rule out designs which are supported by this "punishment-by-beliefs" scenario, we ultimately restrict attention to sequential equilibrium profiles f that maximize (3) given N. We call such a profile f^* *optimal in design* N. Let $W(N)$ denote the value of (3) in the optimal equilibrium f^*.[12] A design N *optimal* if it maximizes W.

[9] Only pure strategies are considered in this analysis.

[10] There are $m(m-1)/2$ possible linked pairs, and $2c$ is the maximal total payoff for each link.

[11] Specifically, suppose there are two designs, N and N'. In design N all residents choose D (the uncooperative equilibrium), even though other equilibria exist. In design N', residents choose an equilibrium that admits some social cooperation even though the uncooperative equilibrium always exists. While design N' then yields a higher value according to (3), the comparison is biased by the selective choice of equilibrium in each.

[12] An explicit implementation approach elicits the value $W(N)$. We omit details as they involve standard arguments involving the use of the Revelation Principle.

Second, for tractability we consider solutions to (3) that ignore integer and remainder problems. For example, suppose that the optimal design called for a replication of finite, component graphs of size r. To implement this design would require a population size divisible by r. For our purposes we simply assume such divisibility. While this entails some loss of generality, our solutions will approximate the actual solutions when the population relative to the solution size is large enough or when the solution leaves a small enough remainder.

3 Local Trigger Strategy Equilibria

3.1 Definition

Here we introduce a type of stationary, "trigger strategy" equilibrium for repeated play in graphs. To facilitate our definition, some definitions and notation are needed.

First, for each individual $i \in M$ and each personal history h_i^t let $N_i(h_i^t)$ denote the set of neighbors that have always chosen the cooperative action, "C", in the past, i.e.,

$$N_i(h_i^t) = \{j \in N_i : \; a_{j\tau} = C, \, \forall \, \tau < t\}.$$

Next, a collection, $\mathcal{T} \subset 2^M$, of sets is *comprehensive* if for any nonempty set $S \in \mathcal{T}$, if $S' \supseteq S$, then $S' \in \mathcal{T}$. Comprehensive collections are those that are closed under the taking of supersets of nonempty sets. Let $CH(S)$ denote the *comprehensive hull*, i.e., the collection of sets defined by including all supersets of the set S.

Definition 1 *A* local trigger strategy for individual i in the neighborhood design N *is a strategy f_i satisfying: for each vector δ, there exists a comprehensive collection $\mathcal{T}_i^\delta \subseteq 2^{N_i}$ of subsets of neighbors of i such that for each personal history h_i,*

$$f_i(h_i; \delta) = \begin{cases} C & if \quad N_i(h_i) \in \mathcal{T}_i^\delta \backslash \{\emptyset\} \\ \\ D & otherwise. \end{cases}$$

In words, a local trigger strategy in a neighborhood design is one in which an individual agrees to cooperate if and only if the permanent cooperators in i's neighborhood is "acceptable" according to $\mathcal{T}_i$. Local trigger strategies require each person to bind his behavior to a judicious, though not unique, selection of trustworthy members of his community. Comprehensivity guarantees that reciprocity is (weakly) increasingly likely, the larger is the set of cooperators. It also guarantees that "D" is an absorbing action for each individual. Since $N_i(h_i)$ can never increase over time, once $N_i(h_i) \notin \mathcal{T}_i^\delta \backslash \{\emptyset\}$, no subsequent cooperating set can be included in $\mathcal{T}_i^\delta \backslash \{\emptyset\}$.

Unfortunately, local trigger strategies entail some loss of generality in the design problem. Since nonstationary strategies are not considered, a relatively impatient neighbor plays "D" permanently. The community might be better off if an impatient type could choose "C" every nth period. Time varying strategies such as this require more complex recall. By contrast, local trigger strategies admit a very simple partition of private histories into two states, "cooperative" and "punishment/free riding," the latter being absorbing.

Hereafter, we associate resident i's local trigger strategy with the set $\mathcal{T}_i^\delta$ and refer to it as i's collection of "trigger sets." The collection $\mathcal{T}_i^\delta = \{\emptyset\}$, for example, corresponds to the strategy, "always play D." The trigger set $\mathcal{T}_i^\delta = CH(S)$ for some $S \subset N_i$ corresponds to the strategy, "play C as long as everyone in S continues to play C; play D otherwise." In this case, individual i's cooperation depends on the continued cooperation of everyone in set S.[13] Such a strategy may be contrasted with the trigger set $\mathcal{T}_i^\delta = \{\{j\}, \{k\}, \{j,k\}\}$ which corresponds to: "play C if <u>either</u> j or k continue to play C; play D otherwise." We are now ready to define our equilibrium concept.

Definition 2 *A* local trigger strategy equilibrium (LTSE) in neighborhood design N *is a collection of trigger strategy profiles, $\mathcal{T} = (\mathcal{T}_i^\delta)_{i \in N,\, \delta \in [0,1)^m}$, that comprises, for each δ, a sequential equilibrium. An* optimal LTSE *is an LTSE that maximizes the social welfare criterion (3) in the class of LTSE.*

Clearly, the set of LTSE is nonempty since it always contains the un oo er ti e e ui li rium" $\mathcal{T}_i^\delta = \{ \}$ f r e ch an ea h δ in ve y n ig bo ho d d si n. aturally, we will be interested in optimal local trigger strategy equilibria. The optimal LTSE strategy of Neighbor 1 in both the 3-person cul-de-sac (Figure 1), and the 3-person grid (Figure 4) is described below in terms of the low, intolerant, and tolerant types: $\delta_L, \delta_I, \delta_T$.

(i) An optimal equilibrium in the cul-de-sac.

$$\mathcal{T}_1^\delta = \begin{cases} \{\{2,3\}\} & if\ \delta_i \in \{\delta_T, \delta_I\},\ \forall i, \\ CH(\{2\}) & if\ \ \delta_3 = \delta_L,\ \delta_2 = \delta_T \\ CH(\{3\}) & if\ \ \delta_2 = \delta_L,\ \delta_3 = \delta_T \\ \{\emptyset\} & otherwise. \end{cases}$$

[13] While the supersets of S are also acceptable, only the cooperation of members of S is essential for i's continued cooperation.

(ii) An optimal equilibrium for the grid.

$$\mathcal{T}_1^\delta = \begin{cases} \{\{2,3\}\} & if\ \delta_1 = \delta_I,\ \delta_j \in \{\delta_T, \delta_I\},\ j = 2,3 \\ CH(\{2\}) & if\ \delta_1 = \delta_T,\ \delta_2 \in \{\delta_T, \delta_I\}, \\ CH(\{3\}) & if\ \delta_1 = \delta_T,\ \delta_2 = \delta_L,\ \delta_3 \in \{\delta_T, \delta_I\} \\ \{\emptyset\} & otherwise. \end{cases}$$

Since partial cooperation is preferable to none at all, some in-equilibrium free riding occurs for certain configurations of δ in each design. Notice that in these particular examples, $\mathcal{T}_1^\delta$ can be expressed as a comprehensive hull of some set. Generally, this need not be the case. Again using the three types, suppose in Figure 6 below that Neighbor 3 is tolerant of a single free rider , i.e., $\delta_3 = \delta_T$, while all others are cooperative but intolerant, i.e., $\delta_i = \delta_I$ for all $i \neq 3$.

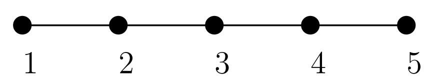


Figure 6

For this configuration of δ, the following is an LTSE: $\mathcal{T}_i^\delta = \{N_i\}$ for each $i \neq 3$, while $\mathcal{T}_3^\delta = \{\{2\}, \{4\}, \{2,4\}\}$. While each of the others requires cooperation from everyone in his neighborhood, Neighbor 3 only requires it from either of his two neighbors. Despite Neighbor 3's failure to punish each neighbor individually, both Neighbors 2 and 4 will cooperate since sufficient deterrence is provided by their end-point neighbors.

3.2 Incentive Constraints

Local trigger strategy equilibria are characterized by two types of incentive constraints. First, if in equilibrium individual i is expected to play "C", then there is no incentive to choose "D" after any history in which only cooperation in one's neighborhood has been observed. A second constraint is necessary to ensure that punishment threats are credible. Player i must find it optimal to punish an individual j who defects to D whenever $N(h_i)\backslash\{j\} \notin \mathcal{T}_i^\delta$.

To characterize these two constraints, fix an LTSE $\mathcal{T}$. Fix a realization δ. For each resident i, fix a subset that possibly includes i: $S_i \subseteq N_i \cup \{i\}$. Let $S = (S_1, \ldots, S_m)$. Define

$$Q^{\mathcal{T}}(S) = \left\{ k \in N : \; S_k \notin \mathcal{T}_k^{\delta} \backslash \{\emptyset\} \right\}$$

so that $Q^{\mathcal{T}}(S)$ denotes the set of all residents $k \in N$ who choose "D" when the set S_k constitutes the neighbors who choose "C". Given an intitial state S^0, the transition law of motion according to $\mathcal{T}$ is given by $S_i^t = S_i^{t-1} \backslash Q^{\mathcal{T}}(S^{t-1})$ for each $t = 1, 2, \ldots$. Here, S_i^t denotes the set of neighbors who choose "C" t periods after initial state S^0. If S^* is a fixed point of the transition map, that is, if

$$S^* = (S_1^*, \ldots, S_m^*) = (\; S_1^* \backslash Q^{\mathcal{T}}(S^*), \ldots, S_m^* \backslash Q^{\mathcal{T}}(S^*) \;),$$

then S_i^* is i's equilibrium set of cooperating neighbors. Observe that $(\emptyset, \ldots, \emptyset)$ is a trivial fixed point of this map. The set of cooperators everywhere is the union denoted by: $\mathcal{S}^* = \cup_i S_i^*$.

Let $\sigma_i^t(S_i) = |S_i^t|$ denoting the number of cooperators in i's neighborhood in the equilibrium continuation, t periods after the initial state $S_i^0 = S_i$. Since the sequence $\{S_i^t\}$ denotes an equilibrium continuation, our definition recursively implies

$$\sigma_i^t(S_i^0) = \sigma_i^{t-1}(S_i^1) = \sigma_i^{t-2}(S_i^2) = \cdots = \sigma_i^1(S_i^{t-1}).$$

It also follows that for each t, $\sigma_i^t(S_i) \geq \sigma_i^t(S_i')$ if $S_i \supseteq S_i'$.

Now fix an LTSE, a resident i, an

To specialize further, suppose that all of i's neighbors are able to punish him if he deviates (as in the standard repeated game). Then (5) reduces to a generalization of the Inequalities (1) and (2) in the Introductory Example of Section 1:

$$\delta_i \geq \min\left\{1, \frac{d-c}{d} + \frac{\ell}{d}\frac{(n_i - s_i^*)}{s_i^*}\right\} \tag{6}$$

As for the second, "perfection" constraint, suppose that S reflects one or more defections from the equilibrium level of cooperation. Then individual i's punishment threat following an observed defection in each set in $\mathcal{T}_i^\delta$ is credible if

$$\begin{aligned} &\sum_{t=0}^{\infty}(1-\delta_i)\delta_i^t\sigma_i^{t+1}(S_i)\ d \\ \geq\ \ &(1-\delta_i)[\sigma_i^1(S_i)c - (n_i - \sigma_i^1(S_i))\ell] + \delta_i\sum_{t=0}^{\infty}(1-\delta_i)\delta_i^t\sigma_i^{t+1}(S_i^1 \cup \{i\})\ d. \end{aligned} \tag{7}$$

In Inequality (7), $i \notin S_i^t$ for all $t \geq 1$ since the equilibrium continuation entails that individual i use his trigger punishment immediately. By comparison, in Inequality (15), $i \in S_i^t$ for each $t < T$ since the equilibrium continuation there entails that i cooperate until date T. One can verify then that (7) is the opposite inequality of (4) for the case when $T = 0$ (i.e., player i chooses "D" in equilibrium immediately in the following period).

Inequality (7) implicitly defines an upper bound, $U(\sigma_i, S_i)$, for i's discount factor. If $\delta_i > U(\sigma_i, S_i)$ then the patient player prefers to tolerate a deviation in his neighborhood for a period. This may occur when, for example, most of his neighbors do not initially observe the deviation. By contrast, $\delta_i \leqslant U(\sigma_i, S_i\backslash\{k\})$ implies that i can credibly punish Neighbor k for defecting from cooperative behavior in state S_i. Hence, $\mathcal{T}_i^\delta$ may be chosen by i to satisfy $S_i\backslash\{k\} \notin \mathcal{T}_i^\delta$. Since local trigger strategi unish k for an initial defection from equilibrium, then he is willing to punish k for defecting when there a fewer than $|S_i^*\backslash\{k\}|$ existing cooperators. That is, $U(\sigma_i, S_i^*\backslash\{k\}) \leqslant U(\sigma_i, S_i')$ for any $S_i' \subseteq S_i^*\backslash\{k\}$. Hence, it suffices to check punishment incentives for single deviations from equilibrium, i.e., whether $\delta_i \leqslant U(\sigma_i, S_i^*\backslash\{k\})$ whenever $S_i^*\backslash\{k\} \notin \mathcal{T}_i^\delta$.

Inequalities (5) and (7) determine incentive constraints required to enforce any LTSE. Specifically, an individual i may cooperate in an LTSE if $\delta_i \geq L(\sigma_i, S_i^*)$, and may punish a defector $k \in S_i^*$ if $\delta_i \leqslant U(\sigma_i, S_i^*\backslash\{k\})$.

4 Optimal Neighborhood Designs

Throughout this section, we assume that optimal LTSE are used (where optimality is restricted in the sense that it is relative to the set of LTSE). The notion of *optimal design* is re at ve o t is la s.

.1 n I it al bs rv ti n

T e r su t b lo re ea s w y h te og ne ty s n ce sa y f r t e p an in pr bl m t be on ri ia . T e n ta io *su p G* ef rs he up or of .

T eo em *1. f s pp* $\subseteq [\frac{-c}{}, 1^{\ m}$*, he an pt ma ne gh or oo de ig is ni ue y g ve by he ax ma cl qu :* $N = M$ $\{i$ *fo al i.*

. I su p G $[0\ \frac{d-}{d})$ *th n a y n ig bo ho d d si n i op im l.*

ro f ( ) L t *s pp* $\subseteq [\frac{-c}{}, 1^{\ m}$. he we ho th t i an cl qu N' it m' $|M|$, he ax ma LT E e ta ls ha ev ry ne la C. ny ef ct on s f ll we by gr m t ig er tr te y i wh ch ve yo e p ay D t er af er If uc an qu li ri m e is s, he $W(\ ';\ ) = m'\ \ 1)$. O se ve ha if $' =$, i e. th cl qu is he ax ma on , t en he la ne 's ri er on s m xi al ve al po si le ei hb rh od es gn . W no sh w t at hi eq il br um xi ts an so he la ne ca re li e $W\ N'\ f)$ $(m\ -1\ c.$ ec ll ha In qu li y ( ) i th re ev nt nc nt ve on tr in wh n $S\ = S_i$. in e t e g im ri ge is la ed y a l i st te $_i^*$ $\{i$, t e i ce ti e c ns ra nt n ( ) r du es o t e I eq al ty n ( ). ut n t e m xi al li ue $S^* = S_j =$ fo al i a d j an so $\frac{1}{i}$ $S^*) =$. T er fo e, he nc nt ve on tr in is $_i \geq \frac{-c}{}$ wh ch learly holds when *supp* $G \subseteq [\frac{d-c}{d}, 1)^m$. Finally, observe that the perfection constraint, (7), never binds in the grim trigger LTSE in cliques since in that case (7) reduces to $0 \geq (n_i - 1) - \ell$.

(2) Clearly the bound $(d-c)/d$ is the minimal lower bound for δ_i in any equilibrium of any neighborhood design. Hence, if *supp* $G \subseteq\ 0, \frac{-c}{})^m$ he in ny ei hb rh od es gn he nl po si le TS is he ne n w ic al in iv du ls la "D ev ry er od Th re or th pl nn r i in iff re t b tw en ll es gn . $\diamondsuit$

he as of bs rv bl , h mo en ou di co nt ac or (i e. di tr bu io G i de en ra e a d p ac s f ll as on he ec or $\hat{\delta}$ . . , $\hat{}$) f r s me δ) s a im le as wh re ne f t e t o h po he es f T eo em al ay ho ds Gi en he es lt th s c se oe no gi e r se o a in er st ng es gn ro le . N r d co di io s u de wh ch he ol Th or m a pl es Li it d i te ac io is os ib y o ti al nl wh n i re on il ab e s ci l c nfl ct ar se Su se ue t a al si wi l a lo fo th s p ss bi it by ss mi g a up or th t i te se ts ot in er al .

4.2 The Full Information Optimal Design

Suppose that the planner knows the exact value of δ, that is, let G be the degenerate distribution that places probability one on some vector $\delta = (\delta_1, ..., \delta_m)$. Define the set - $= \{i : \delta_i \geq \frac{d-c}{d}\}$. Let |- | $= \omega$, which is the number of individuals that are above the lower bound for cooperation. A preliminary result establishes that in the optimal equilibrium in the optimal design, the individuals in - are those that cooperate.

Lemma *If G is a degenerate distribution that places full mass on some δ, then the optimal LTSE in any optimal design satisfies:* $\mathcal{T}_i^\delta \neq \{\emptyset\}$, $\forall i \in$ - .

Note: by the stationarity of LTSE, $\mathcal{T}_i^\delta \neq \{\emptyset\}$ implies that i cooperates in equilibrium.

Proof: Suppose that $N = (N_1, ..., N_m)$ is an optimal design. Suppose, by contradiction, that $a_i^t = D$ for some i, t in the optimal LTSE. Because LTSE are stationary, we must have $a_i^t = D$ for all t. This stationarity, coupled with fact that G is degenerate, allows us to express (3) as

$$W(N) = \frac{1}{m}(1-\delta)\sum_t \left(\sum_{j \in M \setminus \{i\}} \delta_j^t u_j(\widetilde{a}_{N_j}^t(f)) + \delta_i^t u_i(\widetilde{a}_{N_i}^t(f)) \right)$$

Let $\mathcal{S}$ denote the set of cooperators in equilibrium in design N. Let $K = M \setminus (\mathcal{S} \cup \{i\})$ and define $N' = (N_1', ..., N_m')$ by:

$$N_j' = \begin{cases} N_j \setminus \{i\} \text{ if } j \in K \\ N_j \cup \{i\} \text{ if } j \in \mathcal{S} \\ N_i \setminus K \text{ if } j = i \end{cases}$$

In words, N' connects i to all the current cooperators in M, and removes him from all current uncooperative individuals. We now modify the old LTSE as follows: let

$$\mathcal{T}_j^{\delta\prime} = \begin{cases} CH(\mathcal{T}_j^\delta \cup \{\{i\}\}) \text{ if } j \in \mathcal{S} \\ \mathcal{T}_j^\delta \text{ if } j \in K \\ CH(\mathcal{T}_i^\delta \cup \{\mathcal{S}\}) \text{ if } j = i \end{cases}$$

We assert that this is indeed an LTSE and is optimal. First, behavior in K is unchanged. Second, current cooperators in $\mathcal{S}$ continue to cooperate with additional incentives since the same set of punishers is strictly larger as it now includes i. It remains to show that i has an incentive to cooperate. Since he is only connected to those in $\mathcal{S}$ and since $i \in$ - , $\delta_i \geq L(\sigma, \mathcal{S}) = \frac{d-c}{d}$. That is, his incentive contrain ho ds To er fy he er ec io con tr in s o th se ho ou d p ni h i obse ve ha if de ec s, ve yo e p ni he hi . H nc , $\sigma\ (S_i \setminus\ i\}\ = 0$ Th re or , $\delta\ \leqslant U\ \sigma,\ \ \setminus\{\ \})\ \ \ 1$.

o ver fy ha $\mathcal{T}'$ s o ti al o se ve ha in iv du l i cooperates and is connected to all current cooperators in $\mathcal{S}$. He is also disconnected to all free riders. Hence, in every new link created in the new neighborhood structure N' there is mutual cooperation, while in every deleted link, there had been none. Since $\mathcal{T}$ was asserted to be optimal in N, the LTSE $\mathcal{T}'$ is optimal in N'.

We now show that $W(N') > W(N)$. To see this, observe first that $u_j(\widetilde{a}^t_{N'_j}(f)) = u_j(\widetilde{a}^t_{N_j}(f))\ \ \forall j \in K$ since all the defecting individuals who were previously connected to i received a payoff of 0 from interacting with a mutually defective individual. Observe next $u_k(\widetilde{a}^t_{N'_k}(f)) = u_k(\widetilde{a}^t_{N_k}(f)) + c + \ell\ \ \ \forall k \in \mathcal{S}$. Finally, $u_i(\widetilde{a}^t_{N'_i}(f)) = |S^t_i|c$, whereas $u_i(\widetilde{a}^t_{N_i}(f)) = |S^t_i|(d)$. Then, $W(N') = W(N) + |\mathcal{S}|(c+c+\ell-d) = W(N) + |\mathcal{S}|(2c-(d-\ell))$. Since $2c > d-\ell$, $W(N') > W(N)$. This contradicts the assumption that N was an optimal design. $\diamondsuit$

Since the individuals in $M\setminus$- will always choose "D" in any design, the Lemma implies that there are precisely ω cooperators and $m-\omega$ uncooperative individuals. To see how they are all connected in the optimal design, we begin by partitioning the cooperators in - by their degree of tolerance for free riders.

$$
\begin{aligned}
-_{1} &= \{i \in M:\ \delta_i \geq \tfrac{d-c}{d} + \tfrac{\ell}{d}\tfrac{1}{\omega}\} \\
&\vdots \\
-_{k} &= \{i \in M:\ \delta_i \geq \tfrac{d-c}{d} + \tfrac{\ell}{d}\tfrac{k}{\omega}\} \\
&\vdots \\
-_{m-\omega} &= \{i \in M:\ \delta_i \geq \tfrac{d-c}{d} + \tfrac{\ell}{d}\tfrac{m-\omega}{\omega}\}
\end{aligned}
$$

Using Inequality (6), the set $-_k$ is the subset of cooperators in $-$ who are able to tolerate being connected to at most k free riders out of $\omega + k$ total neighbors. Note that $-_k \subset -_{k-1}$, and it may be the case that $-_k = \emptyset$ for some large enough index k. We can now de cr be the full information optimum.

Theorem 2 *Suppose that G is degenerate. Then every optimal neighborhood design satisfies,*

(i) $i \sim j$, $\forall i, j \in -$, *and*

(ii) there exists an ordering, $i_1, i_2, \ldots, i_{m-\omega}$, *of the uncooperative individuals, such that for each* $k = 1, \ldots, m - \omega$, $i_k \sim j$ *iff* $j \in -_k$.

Since uncooperative individuals in $M \backslash -$ are indistinguishable, all optimal designs are equivalent up to payoff-irrelevant permutations of uncooperative individuals. According to the result, the optimal design exhibits a spatial pattern with a cooperative "core" and an uncooperative "fringe" connected to the more tolerant elements of the core. Since the planner has full information, he knows where the social frictions lie. While all pairs of cooperative individuals are linked, each uncooperative individual is connected to as many cooperators as will tolerate his free riding. The gain to each uncooperative neighbor net of the loss to the cooperator is $d - \ell$.

As for the proof, Part (i) in the theorem is a straightforward consequence of the Lemma. Since connectivity among cooperators increases social welfare, it is clear from the Lemma that all cooperators in $-$ should be connected to each other. Part (ii) follows from the assumption that $d - \ell > 0$. When $d - \ell > 0$, there is a net social gain to connecting a free rider to a cooperator provided that the cooperator is not induced to change his behavior as a result. Part (ii) is an iterative application of the equilibrium incentive constrai t ( ). The degree of slackness in the incentive constraints of cooperators determines the feasibility of adding a marginal free rider.[14] However, each additional free rider raises the theshold required to tolerate subsequent free riders. Hence, if $-_k = \emptyset$, then at most only $k - 1$ free riders are tolerated in the network. The rest remain unconnected to cooperators.

The graphs in Figure 7 illustrate three examples of optimal designs. The types L, I and T correspond to the familiar three types δ_L, δ_I and δ_T in Section 1. Recall that low types L never cooperate; intolerant types I only cooperate if surrounded by other cooperators; tolerant types T tolerate a single free rider. Type H is assumed to lie in $-_2$ so that this type tolerates 2 free riders.

[14] This aspect is reminiscent of Bernheim and Whinston's (1990) study of collusion in multi-market firms. The difference is that their design decision is decentralized and limited to one firm's connections.

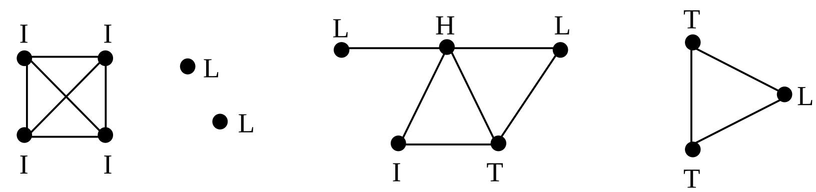


Figure 7:

4.3 Optimal Design with Independent Types

All results in this subsection assume that the discount factors are distributed iid. A simplifying abuse of notation allows that $G_i = G$ for all i. The main result for this Subsection asserts that only cliques are optimal. Hence, the question of structure reduces to one of size.

Theorem 3 *If discount factors are iid, then the optimal neighborhood design partitions M into identically sized cliques.*

The Proof, which is relegated to the Appendix, is long and tedious. The basic idea, however, is not complicated. As it turns out, it is always possible to construct a clique that weakly dominates a given design. To see why, fix an arbitary design. Given this design, there is a link, $i \sim j$ whose average payoff is maximal among those in the design. Let i denote the most likely cooperator of the two in the link. If it is the case that for every δ, a defection by this individual would be punished immediately by all his neighbors, then Neighbor i's "perfection" incentive constraint never binds, i.e., $U(\sigma_i, S_i) = 1$ for all S. This would have been the case had i belonged to a clique. When the incentive constraint does not bind, the conditional probability that an individual is patient enough to cooperate reaches its upper bound. Individual i's cooperation then depends only on the equilibrium proportion cooperators in i's neighborhood. With independent types, this probability does not vary with the rest of the design. Therefore, one can replicate the payoff of the most likely cooperator in the maximal link by simply replicating his neighborhood size throughout the graph. To do this in a way that requires the fewest individuals in a component, all individuals in the component must be maximally connected.

The result establishes as optimal the formation of completely connected coalitions or communities. One person's decision affects all in his local community. In this sense, social cooperation becomes a local public good when types are uncorrelated. The size of each community depends on the unconditional likelihood of each person's cooperation.

Despite positive scale effects of cooperation, the size of these communities need not increase with the scale of society. To see this, consider a clique of size m. Recall that

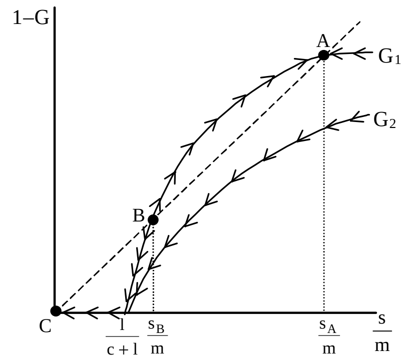


Figure 8:

s^* denotes the number of cooperators in the optimal LTSE. In a clique, the "perfection" constraint never binds, and so everyone punishes a defector. Therefore, the only relevant incentive constraint is Inequality (6). Utilizing this constraint, the conditional probability that an individual cooperates in the optimal LTSE is $1 - G\left(\min\{1, \frac{d-c}{d} + \frac{\ell}{d}\frac{(m-s^*)}{s^*}\}\right)$. A Law of Large Numbers argument asserts that this probability is approximated by the empirical frequency, s^*/m, of cooperation when m is large. Moreover, since s^* takes on values from a finite set, a simple perturbation argument restricts the approximation to stable solutions of the dynamical system $1 - G\left(\min\{1, \frac{d-c}{d} + \frac{\ell}{d}\frac{(m-s_t)}{s_t}\}\right) = s_{t+1}/m$.

Figure 8 exhibits the solutions to this dynamical system for two distributions, G_1 and G_2. Point A is an interior stable fixed point for distribution G_1. If the clique size m is sufficiently large, then s_A^* individuals cooperate. Point C, at which there is no cooperation, is also stable. Point B is also a fixed point of G_1, but is unstable. Indeed with integer values, it is unlikely that s_B^* will be achieved. For distribution G_2, Point C is the only fixed point and is stable. Distributions such as G_2 put limits on the size of each fully connected community.

Theorem 4 *Suppose that discount factors are iid, , and that the only fixed point of*

$$1 - G\left(\min\left\{1, \frac{d-c}{d} + \frac{\ell}{d}\frac{(1-\alpha)}{\alpha}\right\}\right) = \alpha \tag{8}$$

is $\alpha = 0$. Then there is some integer m^* such that any clique size larger than m^* cannot be part of an optimal design.

Proof Let $s(q)$ denote the equilibrium number of cooperators in the optimal LTSE in a clique of size q. A necessary condition for each i to cooperate is $s(q)c - (q - s(q))\ell \geq 0$ or

$s(q)/q \geq \frac{\ell}{c+\ell}$. The Strong Law of Large Numbers implies $s^*(q)/q \to \alpha^*$ a.e., as $q \to \infty$ where α^* satisfies (8). By the hypothesis of the Theorem, $\alpha^* = 0$. Now fix $\epsilon > 0$ such that $\epsilon < \frac{\ell}{c+\ell}$. Given ϵ, there is some m^* such that if $q > m^*$, $s^*(q)/q < \epsilon < \frac{\ell}{c+\ell}$, so that no individual i will cooperate. Hence, the optimal LTSE in the q-clique yields an expected payoff of 0 in the aggregate. Since a 2-person clique has an expected payoff of $\left(1 - G\left(\frac{d-c}{d}\right)\right)^2 c > 0$, the q-clique is suboptimal. $\diamondsuit$

It is not difficult to show that the size threshold m^* weakly decreases in $\ell/(c+\ell)$. That is, optimal clique size diminishes as the relative loss suffered from uncooperative behavior increases. We now identify a class of distributions for which the optimal community size may be easily computed.

Theorem 5 *Given iid types, let supp* $G \subseteq [0, \frac{d-c}{d} + \frac{\ell}{d(m-1)}\ )$ *such that* $G(\frac{d-c}{d} + \frac{\ell}{d(m-1)}) - G(\frac{d-c}{d}) > 0$. *Then the optimal design is a partition into cliques of size q where q is the integer that most closely approximates*

$$1 + \frac{1}{log\left(\frac{1}{1-G(\frac{d-c}{d})}\right)}. \tag{9}$$

Proof If $\frac{d-c}{d} < \delta_i < \frac{d-c}{d} + \frac{\ell}{d(m-1)}$, then i is cooperative but intolerant. That is, i cannot tolerate any uncooperative types. Therefore i only plays C if all n_i neighbors do so. But this is true for every i, and so each indvidual's payoff is given by $\left(1 - G(\frac{d-c}{d})\right)^{n_i - 1}(n_i - 1)c$. Since all individuals are ex ante identical to the planner, the planner will set $n_i = n_j$, choosing to partition M into disjoint neighborhoods, each of which is an identically sized clique that solves

$$\max_q \left(1 - G\left(\frac{d-c}{d}\right)\right)^q (q-1)c \tag{10}$$

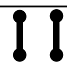
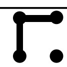
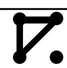
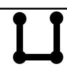
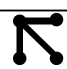
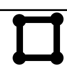
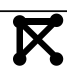
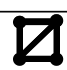
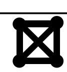
The solution to (10) is the closest integer approximation to (9). $\diamondsuit$

Since the exponential decrease in likelihood of cooperation is of larger order than the returns to scale in community size, the support restriction of G in Theorem 5 suggests strong decreasing returns to community size at some level when partial cooperation is not possible. From this, it follows that for any $2 \leqslant q \leqslant m$, there exist parameters, d, c and ℓ and distribution G such that the optimal design is a clique of size q.

4.4 Correlated Types

When the distribution G admits correlation, then incomplete graphs with low neighbor overlap may be optimal for certain regions of the parameter set. To see this, we work out an example with four people and four types. As before, a low type, δ_L, is below the minimal threshold for cooperation and so will always defect. An "intolerant" type, δ_I, will cooperate so long as every one of his neighbors cooperates and defect if any of his neighbors defects. The "tolerant" type, δ_T, is associated with residents who can withstand one defector for every cooperator, but not more. Finally, highly tolerant types, δ_H, are those that cooperate as long as a single neighbor cooperates.

First, consider a distribution G which puts full mass on the permuations from the set $\{\delta_L, \delta_I, \delta_T, \delta_T\}$. The planner knows there are two tolerant types, one intolerant type, and one uncooperative type (but does not know who is which). The payoffs associated with each of the structures are listed under Case 1 of Table 1 below.

Graph Number	Neighborhood Design	Case 1 $\{\delta_L, \delta_I, \delta_T, \delta_T\}$	Case 2 $\{\delta_L, \delta_H, \delta_H, \delta_H\}$
1		$\frac{1}{2}c$	$\frac{1}{2}c$
2		$\frac{1}{2}c + \frac{1}{6}(d-\ell)$	$\frac{2}{3}c + \frac{1}{6}(d-\ell)$
3		$\frac{2}{3}c + \frac{1}{6}(d-\ell)$	$c + \frac{1}{2}(d-\ell)$
4		$\frac{7}{12}c + \frac{5}{24}(d-\ell)$	$\frac{3}{4}c + \frac{1}{4}(d-\ell)$
5		$\frac{1}{2}c + \frac{1}{8}(d-\ell)$	$\frac{3}{4}c + \frac{3}{16}(d-\ell)$
6		$\frac{2}{3}c + \frac{1}{2}(d-\ell)$	$c + \frac{1}{2}(d-\ell)$
7		$\frac{2}{3}c + \frac{1}{3}(d-\ell)$	$c + \frac{7}{16}(d-\ell)$
8		$\frac{1}{4}c + \frac{1}{12}(d-\ell)$	$\frac{5}{4}c + \frac{5}{8}(d-\ell)$
9		0	$\frac{3}{2}c + \frac{3}{4}(d-\ell)$

To calculate a particular example, consider graph number 4 in Case 1 of Table 1. With probability $\frac{1}{2}$, the low type lies at one of the end points, in which case there is a $\frac{2}{3}$ probability that the low type's neighbor is a tolerant type. This gives an average payoff of $\frac{1}{4}(c + c + 2c + (d-\ell)) = \left(c + \frac{1}{4}(d-\ell)\right)$. There is also a $\frac{1}{2}$ probability that the low type lies at one of the end points and a $\frac{1}{3}$ chance that the low type's neighbor is an intolerant type, yielding an average payoff of $\frac{1}{4}(c + c + (d-\ell)) = \left(\frac{1}{2}c + \frac{1}{4}(d-\ell)\right)$. Finally, there is a $\frac{1}{2}$ probability that the low type is on the interior with a $\frac{2}{3}$ probability that its interior neighbor is a tolerant type. This gives an average payoff of $\frac{1}{4}(c + c + (d-\ell)) = \left(\frac{1}{2}c + \frac{1}{4}(d-\ell)\right)$.

All other possibilities lead to no cooperation, hence a payoff of zero. Overall, the expected average payoff is $\frac{1}{2}\frac{2}{3}\left(c+\frac{1}{4}(d-\ell)\right)+\frac{1}{2}\frac{1}{3}\left(\frac{1}{2}c+\frac{1}{4}(d-\ell)\right)+\frac{1}{2}\frac{2}{3}\left(\frac{1}{2}+\frac{1}{4}(d-\ell)\right)+\frac{1}{2}\frac{1}{3}0=\frac{7}{12}c+\frac{5}{24}(d-\ell)$.

Note that in the clique (Graph 9), a "cascade" effect prevents any cooperation. The low type is, by definition, uncooperative which induces the intolerant type to be uncooperative which induces, in turn, the remaining two types to be uncooperative as well, so that no cooperation is possible. Instead, it turns out that the square or closed grid structure (Graph 6) is optimal. The grid balances the connectivity gains when cooperators are linked against conectivity losses when social conflicts, say between types L and I arise.

Now suppose that G put unit mass on the permuations from the set $\{\delta_L, \delta_H, \delta_H, \delta_H\}$. The resulting expected payoffs are listed under Case 2 in Table 1. Since in this case there are more high types in society, payoffs are, in general, higher across the board. However, along with the change in payoffs, one sees that the optimal neighborhood design has changed. Here, the clique is optimal.

5 Literature and Extensions

A large literature models the determinants of group size, the effects of congestion, peer effects, and the quality and quantity of public services in local jurisdictions. A small sample includes de Bartolome (1990), Benabou (1993), Conley and Wooders (1998), Epple and Romano (1995), Glomm and Lagunoff (1998, 1999), Oates and Schwab (1991), and Scotchmer (1985) to name only a few.[15]

The structure of residential developments is a common and interesting local interaction design problem. Social norms of cooperation differ markedly across different communities. In many neighborhoods, for example, cooperative arrangements in supervision of children, in maintaining communal spaces, or in monitoring the safety of the neighborhood are common. In others less so. Typically, cooperative arrangements, when they exist, are not coerced but instead rely on reciprocity and voluntary "good will" of the residents. Detailed discussion, examples, and surveys of the effect of spatial structure on these norms may be found in Logan and Molotch (1987), Landon (1994), Southworth and Ben-Joseph (1997), and White (1980). Landon, in particular, summarized the relative costs and benefits of cul-de-sacs compared to traditional street grids this way:

> "In especially successful cul-de-sacs, families get to know one another well. Some cul-de-sacs encourage the development of neighborhood events or stimulate daily social routines... [Yet], on a straight block of houses, people may see what the

[15] See Conley and Wooders (1998) for further references.

> neighbors across the street are doing, but people do not feel so tightly entwined with one another... The cul-de-sac often seems to promote more contact than is desired... (pp 42-43)

The present work proposes a tractable way of determining when social contact is likely to be desireable. Problems with local externalities are cast here as mechanism design problems in repeated games. While we are not aware of other work that does this, the nature of the linkage relates the present paper to a growing literature on network externalities. These include a wide variety of applications ranging from macroeconomic growth, e.g., Durlauf (1993), and buyer-supplier relationships, e.g., Kranton and Minehart (1996, 1997), to information transmission, e.g., Bala and Goyal (1998), and social networks, e.g., Chwe (1998).[16] Of particular relevance for the present paper is a subset of this literature which examines strategic (noncoooperative) behavior in networks.[17] Often referred to as *local interaction models*, examples include Anderlini and Ianni (1995), Bala and Goyal (1999), Blume (1993, 1995), Jackson and Wolinsky (1996), Lagunoff and Schreft (1998, 1999), and Morris (1997), among others.[18] Analysis of local interaction applied specifically to Prisoner's Dilemma include Epstein (1998) and Tieman, Houba, and van der Laan (1998).

With a few exceptions, these models tend to be either static or assume adaptive adjustment dynamics. By contrast, the present paper studies repeated game effects in graphs with forward-looking agents. In this dimension, a closer analogue may be found in the study of collusion with multi-market firms (see Bernheim and Whinston (1990)), and in the study of multilateral tariff cooperation (see Bagwell and Staiger (1999) and sources contained therin). Forward looking behavior has also been studied in population games where individuals are repeatedly and randomly paired. See Kandori (1991), Ellison (1994), and Okuno-Fujiwara and Postlewaite (1995). In population games, full cooperation is shown to be supported in repeated Prisoner's Dilemma games when discount factors are high enough.[19]

In addition to local interaction, the present paper introduces heterogeneity in rates of time preference. It is precisely this heterogeneity which makes the linkage design problem nontrivial.[20] This type of heterogeneity has been examined by Harrington (1989) who studies the effect of discount rate differences on collusion in oligoplies, and by Lehrer and Pauzner (1999) who examine heterogeneity in general two-player repeated games.

A number of assumptions are used in the analysis for tractability. Three modifications in future work would enrich the present analysis. First, the use of space and distance would

[16]See also Katz and Shapiro (1994), and Sharkey (1993), and references contained therein.

[17]This is as distinct from *cooperative* game theoretic models of networks, some references for which can be found in Sharkey (1993).

[18]The Morris paper is a good source for further references.

[19]Kandori proves a Folk Theorem property for certain Prisoner's Dilemma stage games. Ellison extends the Kandori result to all PD games when a public randomizing device is available.

[20]Ellison (1994) introduces time preference heterogeneity as an extension. Since the emphasis there on whether or not full cooperation exists, heterogeneity does not play a substantial role.

add a dimension of realism, particularly when the model applies to residential neighborhood interaction. If, for example, externalities diminish with distance, then a planner may prefer to mitigate the consequences of certain neighbor's actions by increasing space between him and others, rather than exclude the neighbor altogether.

Second, the perfect link between information flows and externalities could be severed. The more interesting case occurs when the externalities flow beyond one's observations.[21] In such a case, the moral hazard problem is more extreme in larger networks. This suggests a smaller scale of linkage than before will be preferred by the planner.

Third, the optimal design problem should be re-examined when nonstationary sequential equilibria are considered. Local trigger strategies entail loss of generality since occasional cooperation is not admitted. The incentive constraints can be relaxed if, say, a relatively impatient type need only choose "C" every nth period.

6 Appendix

Proof of Theorem 3

We first define some notation that will be used in the proof:

Fix a neighborhood structure N. Fix $S = (S_1, \ldots, S_m)$ in N. As before, let $\mathcal{S} = \cup_i S_i$. For each i let $s_i = |S_i|$ denote the number of cooperators in i's neighborhood N_i. Then, define

$$G_{n_i s_i} = G_i \left(\min \left\{ 1, \frac{d-c}{d} + \frac{\ell}{d} \frac{(n_i - s_i)}{s_i} \right\} \right)$$

In words, $G_{n_i s_i}$ is the probability that i is too impatient to cooperate when s_i of his neighbors are cooperative. We use the convention that $G_{ms} = G_{mm}$ whenever $s > m$. Observe that $G_{\cdot s_i}$ is increasing in the size n_i of one's neighborhood.

Lemma 1 *Suppose the marginal distributions, G_i, $i \in M$ are independent and identical. Then for any neighborhood design N and any LTSE, the probability that an individual i cooperates, conditional on s_i, is no greater than $1 - G_{n_i s_i}$.*

We refer to $1 - G_{n_i s_i}$ as the upper bound on i's cooperation likelihood.

[21] The case where a player's information extends beyond his externalities is less interesting. To take an extreme case, if residents could condition punishment on the behavior of everyone, then standard repeated game arguments can be used to prove a Folk Theorem when the δ realizations are large enough. Kandori (1992) makes this observation in a random matching model.

Proof Fix any neighborhood design (possibly not optimal). Suppose that an individual i who cooperates in equilibrium would be punished by <u>all</u> others in N_i should he defect at any stage. Of course, an LTSE in which this is true may not exist. However, we need only establish that $1 - G_{n_i s_i}$ is an upper bound, and so if it does exist, then the perfection incentive constraint never binds for any i who cooperates in equilibrium. The upper bound for δ_i is, therefore, $\delta_i \leqslant 1$. The incentive constraint for equilibrium cooperation is then given by

$$s_i c - (n_i - s_i)\ell \geq (1 - \delta_i) s_i d. \tag{11}$$

Note that since all individuals in N_i punish i immediately, the right hand side of (11) is minimal. This, then gives the lowest possible bound for δ_i which is expressed by

$$\delta_i \geq \frac{d - c}{d} + \frac{\ell}{d} \frac{(n_i - s_i)}{s_i}. \tag{12}$$

Since G_i independently distributes δ_i, (12) describes the an upper bound for i's probability of cooperation conditioned on s_i. $\diamondsuit$

Let $\theta : K \to K$ denote a permutation of a subset $K \subseteq M$ of residents. Resident i is re-indexed to be $\theta(i)$ under θ. Let $\Theta(K)$ denote the set of all permutations on the set K. Note that if $|K| = k$, then $|\Theta(K)| = k!$. Denote the subset of i's neighbors who are successors under θ by

$$R_i^\theta = \{ j \in N_i : \ \theta(j) > \theta(i) \}.$$

Let $s_i^\theta = |S_i \cup R_i^\theta|$. Next define

$$G_{n_i s_i}^\theta = G\left(\min\left\{ 1, \frac{d - c}{d} + \frac{\ell}{d} \frac{(n_i - s_i^\theta)}{(s_i^\theta)} \right\} \right)$$

Note that if $\theta(i) = k$, i.e., θ orders i last in the queue, then $R_i^\theta = \emptyset$ and so $G_{n_i s_i}^\theta = G_{n_i s_i}$.

Lemma 2

(i) For each i the upper bound on conditional cooperation likelihood, $1 - G_{n_i s_i}$, is binding if N is a clique.

(ii) If, in any neighborhood structure the upper bound $1 - G_{n_i s_i}$ binds, then the probability, $P(S \big| N)$, that S occurs given the neighborhood structure N is

$$\prod_{i \in \mathcal{S}} (1 - G_{n_i s_i}) \sum_{\Theta(M \setminus \mathcal{S})} \prod_{i \in M \setminus \mathcal{S}} G_{n_i s_i}^\theta \tag{13}$$

Proof (i) Let N be a clique. We show, first, that the upper bound $U(\sigma, S_i)$ determined by the perfection constraint is always unity if N is a clique. To see this, fix an individual i who must decide whether or not to punish a defector $j \in N_i = M$. If all such neighbors choose D following j's defection, either to punish j or because they had previously chosen D, then, using the perfection constraint, i will choose D since

$$(1-\delta_i)(n_i - n_i)d = 0 \geq (n_i - n_i)c - n_i\ell = -n_i\ell.$$

Since this inequality is satisfied for all δ, $U(\sigma, S_i) = 1$, and so there is no upper bound on discount factors. In equilibrium, then, i cooperates is $\delta_i \geq \frac{d-c}{d} + \frac{\ell}{d}\frac{(n_i - s_i)}{s_i}$.

(ii) Fix S. The probability that all $i \in \mathcal{S}$ cooperate given s_i is, precisely, $\prod_{i \in \mathcal{S}}(1 - G_{n_i s_i})$. However, the probability that all $i \in M \backslash \mathcal{S}$ fail to cooperate is not $\prod_{i \in M\backslash \mathcal{S}} G_{n_i s_i}$. The reason is that S cannot be a conditioning event for all individuals simultaneously. Linkage induces correlation in the likelihood that individuals cooperate. Hence, we construct an unconditional probability that "deviant" set $M\backslash\mathcal{S}$ fails to cooperate in N.

Fix an arbitary permutation $\theta \in \Theta(M\backslash\mathcal{S})$. For i such that $\theta(i) = 1$ (first in the queue), $G^{\theta}_{n_i s_i} = G_{n_i n_i} = G(\frac{d-c}{d})$. Hence, the first in the queue under θ fails to cooperate with minimal unconditional probability $G(\frac{d-c}{d})$. Let j be the first individual in the queue who has a neighbor earlier in the queue under θ. Then j fails to cooperate, conditional on an earlier failure, with probability

$$G^{\theta}_{n_j s_j} = G_{n_j(n_j-1)} = G(\frac{d-c}{d} + \frac{\ell}{d}\frac{1}{(n_j - 1)}),$$

and so on. Summing over all permutations gives $\sum_{\Theta(M\backslash\mathcal{S})} \prod_{i \in M\backslash\mathcal{S}} G^{\theta}_{n_i s_i}$. Expression (13) therefore describes $P(S\big|N)$ when each individual's cooperation likelihood is maximal. $\diamondsuit$

We denote a link in the neighborhood structure (i.e., graph) N by $\mathcal{L}_{ij}$. Abusing notation somewhat, we will write $\mathcal{L}_{ij} \in N$ whenever $i \sim j$ in N. Let $V(\mathcal{L}_{ij}; N)$ denote the per person, expected payoff to the pair i and j in N (before the realization of δ). Then $V(\mathcal{L}_{ij}; N)$ may be expressed as

$$V(\mathcal{L}_{ij}; N) = \begin{cases} \frac{1}{2}\left(2c\left[\sum_{\{S\colon \{i,j\} \subseteq \mathcal{S}\}} P(S\big|N)\right] + (d-\ell)\left[\sum_{\{S\colon \{i,j\} \not\subseteq \mathcal{S},\ \{i,j\} \not\subseteq M\backslash\mathcal{S}\}} P(S\big|N)\right]\right) & if\ \mathcal{L}_{ij} \in N \\ 0\ otherwise & \end{cases} \tag{14}$$

In expression (14), if both individuals i and j cooperate, i.e., if $\{i, j,\} \subseteq \mathcal{S}$, then the total payoff in the link is $2c$. If, however, only one of them cooperates, i.e., if $\{i, j\} \nsubseteq \mathcal{S}$ and $\{i, j\} \nsubseteq M \backslash \mathcal{S}$, then the total payoff is $d - \ell$. If neither cooperates, then the total payoff is 0. Note that $W(N) = \frac{1}{m} \sum_{\mathcal{L}_{ij} \in N} V(\mathcal{L}_{ij}; N)$.

Now fix $S = (S_1, \ldots, S_m)$ in N with $\mathcal{S} = \cup_i S_i$, and as before, let $P(S \big| N)$ denote the probability that S occurs in the neighborhood structure N. Let

$$\bar{P}(S \Big| N) \;\equiv\; \prod_{i \in \mathcal{S}} (1 - G_{n_i s_i}) \sum_{\Theta(M \backslash \mathcal{S})} \prod_{i \in M \backslash \mathcal{S}} G^{\theta}_{n_i s_i} \tag{15}$$

Correspondingly, we let $\bar{V}(\mathcal{L}_{ij}; N)$ denote expected payoff of link $\mathcal{L}_{ij}$ when $\bar{P}$ replaces P in the definition (14). Note that by Lemma 2, $\bar{P}$ coincides with P when N is a clique. We say that $\bar{P}$ *weakly first order stochastically dominates* P if for any feasible $\hat{S}$,

$$\sum_{S \supseteq \hat{S}} \bar{P}(S \Big| N) \geq \sum_{S \supseteq \hat{S}} P(S \Big| N)$$

A standard property of stochastic dominance is that for any function g on the vector S satisfying $g(S) \geq g(\hat{S})$ whenever $S \supseteq \hat{S}$, $\sum_S \bar{P}(S \Big| N) g(S) \geq \sum_S P(S \Big| N) g(\hat{S})$. Therefore, if $\bar{P}$ stochastically dominates P in the sense above, then $\bar{V}(\mathcal{L}_{ij}; N) \geq V(\mathcal{L}_{ij}; N)$.[22] The following Lemma establishes that this is the case.

Lemma 3 *Given N, $\bar{P}(\cdot \Big| N)$ weakly stochastically dominates $P(\cdot \Big| N)$.*

Proof This is a direct consequence of Lemma 1. In particular, we may write $P(S \Big| N)$ as

$$P(S \Big| N) \;\equiv\; \prod_{i \in \mathcal{S}} \Pr(a_i = C \Big| s_i, N) \sum_{\Theta(M \backslash \mathcal{S})} \prod_{i \in M \backslash \mathcal{S}} \Pr(a_i = D \Big| s_i^{\theta}, N) \tag{16}$$

Since Lemma 1 asserts that $(1 - G_{n_i s_i}) \geq \Pr(a_i = C \Big| s_i, N)$, weak stochastic dominance of $\bar{P}$ over P follows. $\diamondsuit$

[22]To see this, just take g to be

$$g(S) = \begin{cases} 2c & if \quad \{i, j\} \subseteq \mathcal{S} \\ d - \ell & if \quad \{i, j\} \nsubseteq \mathcal{S}, \; \{i, j\} \nsubseteq M \backslash \mathcal{S} \\ 0 & \qquad otherwise \end{cases}$$

Define $\bar{P}^i(S\Big|N)$ by

$$\begin{aligned}\bar{P}^i(S\Big|N) &\equiv \prod_{j\in\mathcal{S}}(1-G_{n_is_i})\sum_{\Theta(M\backslash\mathcal{S})}\prod_{j\in M\backslash\mathcal{S}}G^{\theta}_{n_is_i} \\ &= (1-G_{n_is_i})^{|\mathcal{S}|}\sum_{\Theta(M\backslash\mathcal{S})}(G^{\theta}_{n_is_i})^{|M\backslash\mathcal{S}|}\end{aligned} \tag{17}$$

In words, for each $j \in M$, $\bar{P}^i(S\Big|N)$ replaces $G_{n_js_j}$ with $G_{n_is_i}$ everywhere in expression (17). Intuitively, $\bar{P}^i(S\Big|N)$ is the probability that would occur if all individuals faced the same neighborhood as individual i.

Lemma 4 *For any neighborhood structure N, there exists an individual i^* such that*

$$\begin{aligned}W^*(N) &\equiv \frac{1}{n_{i^*}}\sum_{\mathcal{L}_{i^*k}\in N}\left(c\left[\sum_{\{S:\,\{i^*,k\}\subseteq\mathcal{S}\}}\bar{P}^{i^*}(S\Big|N)\right]+\frac{(d-\ell)}{2}\left[\sum_{\{S:\,\{i^*,k\}\not\subseteq\mathcal{S},\,\{i^*,k\}\not\subseteq M\backslash\mathcal{S}\}}\bar{P}^{i^*}(S\Big|N)\right]\right) \\ &\geq W(N).\end{aligned} \tag{18}$$

Proof First, observe from (17) that for each i, $\bar{P}^i(S\Big|N)$ does not vary with any of i's neighbors k. Therefore, $W^*(N)$ is invariant over all neighbors k, and so $W^*(N)$ reduces to

$$c\left[\sum_{\{S:\,\{i^*,k\}\subseteq\mathcal{S}\}}\bar{P}^{i^*}(S\Big|N)\right]+\frac{(d-\ell)}{2}\left[\sum_{\{S:\,\{i^*,k\}\not\subseteq\mathcal{S},\,\{i^*,k\}\not\subseteq M\backslash\mathcal{S}\}}\bar{P}^{i^*}(S\Big|N)\right]$$

We now find i^* as follows. Let i^* satisfy $i^* = \arg\max \sum_{\{S:\,i\in\mathcal{S}\}}\bar{P}(S\Big|N)$. That is, i^* has the highest likelihood of cooperation. Alternatively, i^* maximizes $E[(1-G_{n_is_i})|N]$. But then, replacing everywhere $1-G_{n_js_j}$ with $1-G_{n_is_i}$ in the definition (17) yields a $\bar{P}^{i^*}$ that weakly stochastically dominates $\bar{P}$. Specifically,

$$\sum_{\{S:\,i\in\mathcal{S}\}}\bar{P}^{i^*}(S\Big|N) \geq \sum_{\{S:\,i\in\mathcal{S}\}}\bar{P}(S\Big|N)$$

Consequently, by Lemma 3,

$$W^*(N) \geq \bar{V}(\mathcal{L}_{ij};N) \geq V(\mathcal{L}_{ij};N)\;\forall\, i,j$$

for all i and j, and so

$$W^*(N) \geq \frac{1}{m} \sum_{\mathcal{L}_{ij} \in N} V(\mathcal{L}_{ij}; N) = W(N)$$

$\diamondsuit$

For the remainder of the proof, let N be any neighborhood structure. We construct a clique N^* that satisfies $W(N^*) \geq W(N)$.

Let N^* denote the unique clique on neighborhood N_{i^*}. That is, $N_i = N_j$ for all $i, j \in N_{i^*} \cup \{i^*\}$. Naturally, this also means that $S_i = S_j$ in the clique. We now show that the clique N^* dominates N.

Observe, first, that in the clique, by Lemma 1, 2 and the symmetry of the optimal LTSE in N^*, $P(S_{i^*} \Big| N^*) = \bar{P}(S_{i^*} \Big| N^*) = \bar{P}^{i^*}(S_{i^*} \Big| N^*)$. To simplify notation, let $n^* = n_{i^*}$ and $s^* = s_{i^*} = |S_{i^*}|$. Observe

$$\begin{aligned}
\sum_{\{S\colon \{i^*,k\} \subseteq \mathcal{S}\}} \bar{P}^{i^*}(S \Big| N) &= \sum_{\{S\colon \{i^*,k\} \subseteq \mathcal{S}\}} (1 - G_{n^*s^*})^{|\mathcal{S}|} \sum_{\Theta(M \setminus \mathcal{S})} (G^{\theta}_{n^*s^*})^{|M \setminus \mathcal{S}|} \\
&= \sum_{\{S_{i^*}\colon \{i^*,k\} \subseteq S_{i^*}\}} \sum_{\{\hat{S}\colon \hat{S}_{i^*} = S_{i^*}\}} (1 - G_{n^*s^*})^{|\hat{\mathcal{S}}|} \sum_{\Theta(M \setminus \hat{\mathcal{S}})} (G^{\theta}_{n^*s^*})^{|M \setminus \hat{\mathcal{S}}|} \\
&= \sum_{\{S_{i^*}\colon \{i^*,k\} \subseteq S_{i^*}\}} (1 - G_{n^*s^*})^{s^*} \sum_{\Theta(N_{i^*} \setminus S_{i^*})} (G^{\theta}_{n^*s^*})^{n^*-s^*} \sum_{\{\hat{S}\colon \hat{S}_{i^*} = S_{i^*}\}} P^{i^*}(\hat{S} \Big| N,\ S_{i^*}) \\
&= \sum_{\{S_{i^*}\colon \{i^*,k\} \subseteq S_{i^*}\}} (1 - G_{n^*s^*})^{s^*} \sum_{\Theta(N_{i^*} \setminus S_{i^*})} (G^{\theta}_{n^*s^*})^{n^*-s^*} \\
&= \sum_{\{S_{i^*}\colon \{i^*,k\} \subseteq S_{i^*}\}} P(S_{i^*} \Big| N^*)
\end{aligned} \tag{19}$$

and, therefore,

$$\sum_{\{S\colon \{i^*,k\} \subseteq \mathcal{S}\}} \bar{P}^{i^*}(S \Big| N) = \sum_{\{S_{i^*}\colon \{i^*,k\} \subseteq S_{i^*}\}} P(S_{i^*} \Big| N^*)$$

Similarly, one can show

$$\sum_{\{S\colon \{i^*,k\} \not\subseteq \mathcal{S},\ \{i^*,k\} \not\subseteq M \setminus \mathcal{S}\}} \bar{P}^{i^*}(S \Big| N) = \sum_{\{S_{i^*}\colon \{i^*,k\} \not\subseteq S_{i^*},\ \{i^*,k\} \not\subseteq N_{i^*} \setminus S_{i^*}\}} P(S_{i^*} \Big| N^*)$$

We now have:

$$
\begin{aligned}
W(N^*) &= \frac{1}{n^*}\sum_{\mathcal{L}_{i^*k}\in N^*}\left(c\left[\sum_{\{S_{i^*}:\,\{i^*,k\}\subseteq\mathcal{S}\}} P(S\big|N^*)\right]+\frac{(d-\ell)}{2}\left[\sum_{\{S_{i^*}:\,\{i^*,k\}\not\subseteq\mathcal{S}\}} P(S\big|N^*)\right]\right)\\
&\geq \frac{1}{n^*}\sum_{\mathcal{L}_{i^*k}\in N}\left(c\left[\sum_{\{S:\,\{i^*,k\}\subseteq\mathcal{S}\}} \bar{P}^{i^*}(S\big|N)\right]+\frac{(d-\ell)}{2}\left[\sum_{\{S:\,\{i^*,k\}\not\subseteq\mathcal{S}\}} \bar{P}^{i^*}(S\big|N)\right]\right)\\
&= W^*(N)\\
&\geq W(N)
\end{aligned}
$$

The first inequality follows from (19) and the fact that N^* is a clique, i.e., $|\{\mathcal{L}_{i^*k}\in N^*\}| = n^* = |\{\mathcal{L}_{i^*k}\in N\}|$. The second equality and second inequality follow from Lemma 4. Thus, we conclude the Theorem. $\diamondsuit$ $\diamondsuit$

References

[1] Anderlini, L. and A. Ianni (1996), "Path Dependence and Learning from Neighbors," *Games and Economic Behavior*, 13: 141-77.

[2] Bagwell, K. and R. Staiger (1999), "Regionalism and Multilateral Tariff Co-operation," in *International Trade Policy and the Pacific Rim*, IEA Conference Volume No. 20, J. Piggot and A. Woodland, Eds., Great Britain: MacMillan.

[3] Bala, V. and S. Goyal (1998), "Learning from Neighbors," *Review of Economic Studies*, 65: 595-621.

[4] Bala, V. and S. Goyal (1999), "A Non-Cooperative Model of Network Formation," *Econometrica*, forthcoming.

[5] Benabou, R. (1993), "Working of a City: Location, Education and Production," *Quarterly Journal of Economics*,

[6] Bernheim, B. D. and M. Whinston (1990), "Multimarket Contact and Collusive Behavior," *Rand Journal of Economics*, 21: 1-26.

[7] Blume, L. (1993), "The Statistical Mechanics of Strategic Interaction," *Games and Economic Behavior*, 5: 387-424.

[8] Blume, L. (1995), "Evolutionary Equilibrium with Forward Looking Players," mimeo.

[9] Conley, J. and M. Wooders (1998), "Tiebout Economies with Differential Genetic Types and Endogenously Chosen Crowding Characteristics," Mimeo.

[10] Chwe, M. (1998), "Communication and Coordination in Social Networks," mimeo, University of Chicago.

[11] De Bartolome (1990), "Equilibrium and Inefficiency in a Community Model with Peer Group Effects," *Journal of Political Economy*, 98: 110-33.

[12] Durlauf, S. (1993), "Nonergodic Economic Growth," *Review of Economic Studies*, 60: 349-66.

[13] Ellison, G. (1994), "Cooperation in the Prisoner's Dilemma with Anonymous Random Matching," *Review of Economic Studies*, 61: 567-88.

[14] Epstein, J. (1998), "Zones of Cooperation in a Demographic Prisoner's Dilemma," *Complexity*, 4: 36-48.

[15] Epple, D. and R. Romano (1995), "Public School Choice and Finance Policies, Neighborhood Formation, and the Distribution of Educational Benefits, mimeo, July.

[16] Fudenberg, D. and E. Maskin (1986), "The Folk Theorem in Repeated Games with Discounting or with Incomplete Information," *Econometrica*, 54: 533-56.

[17] Glomm, G. and R. Lagunoff (1997), "A Tiebout Theory of Public vs Private Provision of Collective Goods," *Journal of Public Economics*, 68: 91-112.

[18] Glomm, G. and R. Lagunoff (1998), "A Dynamic Tiebout Theory of Voluntary vs Involuntary Provision of Public Goods," *Review of Economic Studies*, forthcoming.

[19] Harrington, J. (1989), "Collusion Among Asymmetric Firms: The Case of Different Discount Factors, *International Journal of Industrial Organization*, 7: 289-307.

[20] Jackson, M. and A. Wolinsky (1996), "A Strategic Model of Social and Economic Networks," *Journal of Economic Theory*, 71: 44-74.

[21] Kandori, M. (1992), "Social Norms and Community Enforcement," *Review of Economic Studies*, 59: 63-80.

[22] Katz, M. and C. Shapiro (1994), "Systems Competition and Network Effects," *Journal of Economic Perspectives*, 8: 93-115.

[23] Kranton, R. and D. Minehart (1996), "Buyer-Supplier Networks," mimeo.

[24] Kranton, R. and D. Minehart (1997), "Link Patterns in Buyer-Supplier Networks:Incentives and Allocation in Graphs," mimeo.

[25] Lagunoff, R. and S. Schreft (1998a), "A Model of Financial Fragility," mimeo.

[26] Lagunoff, R. and S. Schreft (1999), "Financial Fragility with Rational and Irrational Exuberance," *Journal of Money, Credit, and Banking*, forthcoming.

[27] Langdon, P. (1994), *A Better Place to Live: Reshaping the American Suburb*, Amherst: University of Massachussetts Press.

[28] Lehrer, E. and A. Pauzner (1999), "Repeated Games with Differential Time Preferences," *Econometrica*, 67: 393-412.

[29] Logan, J. and H. Molotch (1987), *Urban Fortunes: The Political Economy of Place*, Berkeley: The University of California Press.

[30] Morris, S. (1997), "Contagion," mimeo.

[31] Oates, W. and R. Schwab (1991), "Community Composition and the Provision of Local Public Goods," *Journal of Public Economics*, 44: 217-238.

[32] Okuno- uj wa a, . a d A Po tl wa te 19 5) "S ci l N rm in an om at hi g G me ," *am s a d E on mi Be av or* 9: 9- 09

[3] S ot hm r, . (98), Pr fit ax mi in Cl bs " *J ur al f P bl c E on mi s*, 7: 5- 5.

34 Sh rk y, . (99), Ne wo k M de s i Ec no ic ," *an bo k o Op ra io s R se rc an Ma ag me t S ie ce*

[3] S ut wo th M. nd . B n- os ph 19 7) *St ee s a d t e S ap ng f T wn an Ci ie* , N w Y rk Mc ra -H ll

[3] T em n, ., . H ub , a d G va de La n (99), Co pe at on n a ul i- im ns on l L ca In er ct on od l, mi eo Ti be ge In ti ut .

[7] hi e, . (98), *me ic n N ig bo ho ds nd es de ti l D ffe en ia io* , N w Y rk Ru se l S ge ou da io .